Statistics II
Lesson 1. Inference on one population

Year 2009/10

Lesson 1. Inference on one population

Contents

Lesson 1. Inference on one population

Learning goals

- Know how to estimate population values of means, variances and proportions from simple random samples
- Know how to construct confidence intervals for the mean of one population
  - In the case of a normal distribution
  - In the general case for large samples
- Know how to construct confidence intervals for the population proportion from large samples
- Know how to construct confidence intervals for the variance of one normal population

Lesson 1. Inference on one population

Bibliography references

- Meyer, P. "Probabilidad y aplicaciones estadísticas" (1992)
  - Chapter 14
- Newbold, P. "Estadística para los negocios y la economía" (1997)
  - Chapters 7 and 8 (up to 8.6)

Inference

Definitions

- Inference: the process of obtaining information corresponding to unknown population values from sample values
- Parameter: an unknown population value that we wish to approximate using sample values
- Statistic: a function of the information avaliable in the sample
- Estimator: a random variable that depends on sample information and whose value approximates the value of the parameter of interest
- Estimation: a concrete value for the estimator associated to a specific sample

Inference

Example

We wish to estimate the average yearly household food expenditure in a given region from a sample of 200 households

- The parameter of interest would be the average value of the expenditure in the region
- A relevant statistic in this case would be the sum of all expenditures of the households in the sample
- A reasonable estimator would be the average household expenditure in a sample
- If for a given sample the average food expenditure is 3.500 euros, the estimation of the average yearly expenditure in the region would be 3.500 euros.

Point estimation

- Population parameters of interest:
  - mean or variance of a population, or the proportion in the population possesing a given characteristic
- Selecting an estimator:
  - Intuitively: for example, from equivalent values in the sample
  - Or alternatively those estimators having the best properties

Properties of point estimators

- Bias: the difference between the mean of the estimator and the value of the parameter
  - If the parameter of interest is μ and the estimator is $\hat{\mu}$, its bias is defined as

$$\text{Bias}[\hat{\mu}] = \text{E}[\hat{\mu}] - \mu$$

- Unbiased estimators: those having bias equal to zero
  - If the parameter is the population mean μ, the sample mean $\bar{X}$ has zero bias

Point estimation

Properties of point estimators

- Efficiency: the value of the variance for the estimator
  - A measure related to the precision of the estimator
  - An estimator is more efficient than others if its variance is smaller
  - Relative efficiency for two estimators of a given parameter, $\hat{\theta}_1$ and $\hat{\theta}_2$,

$$\text{Relative efficiency} = \frac{\text{Var}[\hat{\theta}_1]}{\text{Var}[\hat{\theta}_2]}$$

Comparing estimators

- Best estimator: a minimum variance unbiased estimator
  - Not always known
- Selection criterion: mean squared error
  - A combination of the two preceding criteria
  - The mean squared error (MSE) of an estimator $\hat{\theta}$ is defined as

$$\text{MSE}[\hat{\theta}] = \text{E}[(\hat{\theta} - \theta)^2] = \text{Var}[\hat{\theta}] + (\text{Bias}[\hat{\theta}])^2$$

Point estimation

Selecting estimators

- Minimum variance unbiased estimators
  - The sample mean for a sample of normal observations
  - The sample variance for a sample of normal observations
  - The sample proportion for a sample of binomial observations
- If an estimator with good properties is not known in advance
  - General procedures to define estimators with reasonable properties
    - Maximum likelihood
    - Method of moments

Point estimation

Exercise 1.1

- From a sample of units of a given product sold in eight days,

$$8 \quad 6 \quad 11 \quad 9 \quad 8 \quad 10 \quad 5 \quad 7$$

  - obtain point estimations for the following population parameters: mean, variance, standard deviation, proportion of days with sales above 7 units
  - if the units sold during another six-day period have been

$$9 \quad 8 \quad 9 \quad 10 \quad 7 \quad 10$$

    compute an estimation for the difference of the means in the units sold during both periods

Results

$$\begin{aligned}
\bar{x} &= (8+6+11+9+8+10+5+7)/8 = 8 \\
s^2 &= \tfrac{1}{n-1}\sum_i (x_i - \bar{x})^2 = 4, \quad s = \sqrt{s^2} = 2 \\
\hat{p} &= (1+0+1+1+1+1+0+0)/8 = 0{,}625 \\
\bar{x} - \bar{y} &= 8 - (9+8+9+10+7+10)/6 = -0{,}833
\end{aligned}$$

Point estimation

Exercise 1.1

- From a sample of units of a given product sold in eight days,

$$8 \quad 6 \quad 11 \quad 9 \quad 8 \quad 10 \quad 5 \quad 7$$

  - obtain point estimations for the following population parameters: mean, variance, standard deviation, proportion of days with sales above 7 units
  - if the units sold during another six-day period have been

$$9 \quad 8 \quad 9 \quad 10 \quad 7 \quad 10$$

    compute an estimation for the difference of the means in the units sold during both periods

Results

$$\begin{aligned}
\bar{x} &= (8+6+11+9+8+10+5+7)/8 = 8 \\
s^2 &= \tfrac{1}{n-1}\sum_i (x_i - \bar{x})^2 = 4, \quad s = \sqrt{s^2} = 2 \\
\hat{p} &= (1+0+1+1+1+1+0+0)/8 = 0{,}625 \\
\bar{x} - \bar{y} &= 8 - (9+8+9+10+7+10)/6 = -0{,}833
\end{aligned}$$

Estimation using confidence intervals

Motivation

- In many practical cases the information corresponding to a point estimate is not enough
  - It is also important to have information related to the error size
  - For example, an estimate of annual growth of 0,5 % would have very different implications if the correct value may vary between 0,3 % and 0,7 %, or if this value may be between -1,5 % and 3,5 %
  - In these cases we may wish to know some information related to the precision of the point estimator
- The most usual way to provide this information is to compute an interval estimator
  - Confidence interval: a range of values that includes the correct value of the parameter of interest with high probability

Estimation using confidence intervals

Concept

- Interval estimator
  - A rule based on sample information
  - That provides an interval containing the correct value of the parameter
  - With high probability
- For a parameter θ, given a value $1-\alpha$ between 0 and 1, the confidence level, an interval estimator is defined as two random variables $\hat{\theta}_A$ and $\hat{\theta}_B$ satisfying
$$P(\hat{\theta}_A \leq \theta \leq \hat{\theta}_B) = 1-\alpha$$
- For two concrete values of these random variables, a and b, we obtain an interval $[a, b]$ that we call a confidence interval at the $100(1-\alpha)\,\%$ level for θ
  - $1-\alpha$ is known as the confidence level of the interval
  - If we generate many pairs a and b using the rule defining the interval estimator, it holds that $\theta \in [a, b]$ for $100(1-\alpha)\,\%$ of the pairs (but not always)

Computing confidence intervals

General comments

- The confidence interval is associated to a given probability, the confidence level
- From the definition of the values defining the interval estimator, $\hat{\theta}_A$ y $\hat{\theta}_B$,

$$P(\hat{\theta}_A \leq \theta \leq \hat{\theta}_B) = 1 - \alpha$$

  these values could be obtained from the values of quantiles corresponding to the distribution of the estimator, $\hat{\theta}$
- We need to know the distribution of a quantity that relates θ and $\hat{\theta}$, in order to compute these quantiles
- This distribution is the basis for the computation of confidence intervals. It depends on
  - The parameter we wish to estimate (mean, variance)
  - The population distribution
  - The information that may be available (for example, if we know the value of other parameters)
- We study in this lesson different particular cases (for different parameters, distributions)

The mean of a normal population with known variance

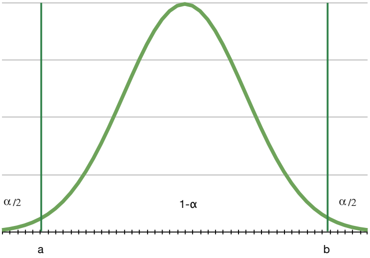
Hypotheses and goal

- We consider first a particularly simple, although not very realistic, case
- We assume that
  - we have a simple random sample of n observations
  - the population follows a normal distribution
  - we know the population variance σ^2
- Goal: construct a confidence interval for the (unknown) population mean μ
  - For a confidence level $1 - \alpha$, either prespecified or selected by us

The mean of a normal population with known variance

Procedure

- Let $X_1, \ldots, X_n$ denote the simple random sample and $\bar{X}$ its sample mean, our point estimator
- Our first step is to obtain information on the distribution of a variable that relates μ and $\bar{X}$
- From this distribution we obtain a pair of values a and b that define the interval (for the variable) having the desired probability
- From these values we define an interval for μ



The mean of a normal population with known variance

Procedure

- For the case under consideration, the distribution of the sample mean satisfies

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

  - A known distribution
  - Relating $\bar{X}$ and μ

- We construct an interval containing the desired probability for a standard normal distribution, finding a value $z_{\alpha/2}$ satisfying

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$

  - $z_{\alpha/2}$ is the value such that a standard normal distribution takes larger values with probability equal to $\alpha/2$

The mean of a normal population with known variance

Procedure

- The following interval has the desired probability

$$-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}$$

- Replacing the sample values and solving for μ we obtain the desired confidence interval

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Computing confidence intervals

Exercise 1.2

A bottling process for a given liquid produces bottles whose weight follows a normal distribution with standard deviation equal to 55 gr. A simple random sample of 50 bottles has been selected; its mean weight has been 980 gr. Compute a confidence interval at 99 % for the mean weight of all bottles from the process

Results

$$\frac{\bar{X}-\mu}{55/\sqrt{50}} \sim N(0,1)$$

$$\alpha = 1-0{,}99 = 0{,}01, \qquad z_{\alpha/2} = z_{0{,}005} = 2{,}576$$

$$-2{,}576 \leq \frac{980-\mu}{55/\sqrt{50}} \leq 2{,}576$$

$$959{,}96 \leq \mu \leq 1000{,}04$$

Computing confidence intervals

Exercise 1.2

A bottling process for a given liquid produces bottles whose weight follows a normal distribution with standard deviation equal to 55 gr. A simple random sample of 50 bottles has been selected; its mean weight has been 980 gr. Compute a confidence interval at 99 % for the mean weight of all bottles from the process

Results

$$\begin{aligned}
\frac{\bar{X}-\mu}{55/\sqrt{50}} &\sim N(0,1) \\
\alpha &= 1-0{,}99 = 0{,}01, \qquad z_{\alpha/2} = z_{0{,}005} = 2{,}576 \\
& -2{,}576 \leq \frac{980-\mu}{55/\sqrt{50}} \leq 2{,}576 \\
& 959{,}96 \leq \mu \leq 1000{,}04
\end{aligned}$$

Computing confidence intervals

General procedure

Steps:

1. Identify the variable with a known distribution and the distribution we will use to construct the confidence interval
2. Find the percentiles of the distribution that correspond to the selected confidence level
3. Construct the interval for the variable with known distribution
4. Replace the sample values in the interval
5. Solve for the value of the parameter in the interval, to obtain another interval, specific for this parameter

Computing confidence intervals

Properties of the interval

- The size of the confidence interval is a measure of the precision in the estimation
- In the preceding case this size was given as

$$\frac{2z_{\alpha/2}\sigma}{\sqrt{n}}$$

- Thus, the precision depended on
  - The standard deviation of the population. The larger it is, the lower the precision in the estimation
  - The sample size. The precision increases as the size increases
  - The confidence level. If we select a higher level we obtain a larger interval

Computing confidence intervals

Exercise 1.3

For the data in exercise 1.2, compute the changes in the confidence interval if

- the sample size increases to 100 (for the same value of the sample mean)
- the confidence level is modified to 95 %

Results

$$-2{,}576 \leq \frac{980-\mu}{55/\sqrt{100}} \leq 2{,}576$$

$$965{,}83 \leq \mu \leq 994{,}17$$

$$\alpha = 1-0{,}95 = 0{,}05, \qquad z_{\alpha/2} = z_{0{,}025} = 1{,}96$$

$$-1{,}96 \leq \frac{980-\mu}{55/\sqrt{50}} \leq 1{,}96$$

$$964{,}75 \leq \mu \leq 995{,}25$$

Computing confidence intervals

Exercise 1.3

For the data in exercise 1.2, compute the changes in the confidence interval if

- the sample size increases to 100 (for the same value of the sample mean)
- the confidence level is modified to 95 %

Results

$$-2{,}576 \leq \frac{980-\mu}{55/\sqrt{100}} \leq 2{,}576$$

$$965{,}83 \leq \mu \leq 994{,}17$$

$$\alpha = 1-0{,}95 = 0{,}05, \qquad z_{\alpha/2} = z_{0{,}025} = 1{,}96$$

$$-1{,}96 \leq \frac{980-\mu}{55/\sqrt{50}} \leq 1{,}96$$

$$964{,}75 \leq \mu \leq 995{,}25$$

The mean of a population from large samples

Motivation

- In many practical cases we do not know if the population distribution is normal or the value of its standard deviation
- If the sample size is sufficiently large, the central limit theorem allows us to construct approximate confidence intervals

Hipotheses and goal

- We assume that
  - we have a simple random sample of size n
  - the size of the sample is sufficiently large
- Goal: construct an approximate confidence interval for the (unknown) population mean μ
  - For a confidence level $1 - \alpha$, either prespecified or selected by us

The mean of a population from large samples

Procedure

- For the case we are considering, the central limit theorem specifies that for n large enough it holds approximately that

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim N(0,1)$$

  where S denotes the sample standard deviation
  - It is the same distribution as in the preceding case
  - It provides a relationship between $\bar{X}$ y μ
- We build as before an interval that contains the desired probability under a standard normal distribution, for a value $z_{\alpha/2}$ satisfying

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$

The mean of a population from large samples

Procedure

- The following interval has the desired probability under the distribution

$$-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq z_{\alpha/2}$$

- Replacing the sample values and solving for the value of μ in the inequalities we obtain the desired confidence interval

$$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}$$

The mean of a population from large samples

Exercise 1.4

A survey has been conducted on 60 persons. Each person provided a rating between 0 and 5 corresponding to their perception of the quality of a given service. The average rating from the sample was 2.8 and the sample standard deviation was 0.7. Compute a confidence interval at the 90 % level for the average rating of the service in the population

Results

$$\alpha = 1 - 0{,}9 = 0{,}1, \qquad z_{\alpha/2} = z_{0{,}05} = 1{,}645$$

$$-1{,}645 \leq \frac{2{,}8 - \mu}{0{,}7/\sqrt{60}} \leq 1{,}645$$

$$2{,}65 \leq \mu \leq 2{,}95$$

The mean of a population from large samples

Exercise 1.4

A survey has been conducted on 60 persons. Each person provided a rating between 0 and 5 corresponding to their perception of the quality of a given service. The average rating from the sample was 2.8 and the sample standard deviation was 0.7. Compute a confidence interval at the 90 % level for the average rating of the service in the population

Results

$$\alpha = 1 - 0{,}9 = 0{,}1, \qquad z_{\alpha/2} = z_{0{,}05} = 1{,}645$$

$$-1{,}645 \leq \frac{2{,}8 - \mu}{0{,}7/\sqrt{60}} \leq 1{,}645$$

$$2{,}65 \leq \mu \leq 2{,}95$$

Proportions from large samples

Motivation

- We wish to estimate the proportion of a population that satisfies a certain condition, from sample data
- Estimating proportions is a particular case of the preceding one when we had nonnormal data
- Our estimator in this case will be the sample proportion
- If X_i represents whether a member of the simple random sample of size n satisfies the condition, or does not satisfy it, and the probability of satisfaction is p, then X_i follows a Bernoulli distribution
- We wish to estimate p, the proportion in the population that satisfies the condition
- Using the sample proportion $\hat{p} = \sum_i X_i / n$
  - $\hat{p}$ is a sample mean

Proportions from large samples

Hypotheses and goal

- We assume that
  - we have a simple random sample of size n, where each observation takes either the value 0 or 1
  - the sample size is sufficiently large
- Goal: construct an approximate confidence interval for the (unknown) population proportion p
  - For a confidence level $1 - \alpha$, either prespecified or selected by us

Proportions from large samples

Procedure

- In our case, compared with the preceding one, $\mu = p$, $\sigma^2 = p(1-p)$
- The central limit theorem states that for large n it holds approximately that

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \sim N(0,1)$$

  where $\hat{p}$ denotes the proportion in the sample
- We approximate $p(1-p)$ with the corresponding sample value $\hat{p}(1-\hat{p})$
  - The resulting variable follows approximately the same distribution
- We construct, as before, an interval contaning the desired probability for a standard normal distribution, computing a value $z_{\alpha/2}$ satisfying

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$

Proportions from large samples

Procedure

- The following interval has the desired probability

$$-z_{\alpha/2} \leq \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \leq z_{\alpha/2}$$

- Replacing sample values and solving for the value of p in the inequalities we obtain the desired confidence interval

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Proportions from large samples

Exercise 1.5

In a sample of 200 patients it has been observed that the number having serious complications associated to a given illness is 38. Compute a confidence interval at the 99 % level for the proportion of patients in the population that may have serious complications associated to that illness

Results

$$\hat{p} = 38/200 = 0{,}19$$

$$\alpha = 1 - 0{,}99 = 0{,}01, \qquad z_{\alpha/2} = z_{0{,}005} = 2{,}576$$

$$-2{,}576 \leq \frac{0{,}19 - p}{\sqrt{0{,}19(1 - 0{,}19)/200}} \leq 2{,}576$$

$$0{,}119 \leq \mu \leq 0{,}261$$

Proportions from large samples

Exercise 1.5

In a sample of 200 patients it has been observed that the number having serious complications associated to a given illness is 38. Compute a confidence interval at the 99 % level for the proportion of patients in the population that may have serious complications associated to that illness

Results

$$\begin{aligned} \hat{p} &= 38/200 = 0{,}19 \\ \alpha &= 1 - 0{,}99 = 0{,}01, \qquad z_{\alpha/2} = z_{0{,}005} = 2{,}576 \\ & -2{,}576 \leq \frac{0{,}19 - p}{\sqrt{0{,}19(1-0{,}19)/200}} \leq 2{,}576 \\ & 0{,}119 \leq \mu \leq 0{,}261 \end{aligned}$$

The mean of a normal population with unknown variance

Motivation

- We wish to estimate the mean of the population
- And we know that the distribution in the population is normal
- But we do not know the variance of the population
- If the sample size is small the preceding results would not be applicable
- But for this particular case we know the distribution of the sample mean for any sample size

Hypotheses and goal

- We assume that
  - we have a simple random sample of size n
  - the population follows a normal distribution
- Goal: construct a confidence interval for the (unknown) population mean μ
  - For a confidence level $1 - \alpha$, either prespecified or selected by us

The mean of a normal population with unknown variance

Procedure

- In this case the basic distribution result is

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

  where S denotes the sample standard deviation and t_{n-1} denotes the Student-t distribution with $n-1$ degrees of freedom
  - A symmetric distribution (around zero) similar to the normal one (it converges to a normal distribution as n increases)
- We again construct an interval containing the desired probability, but we do it for a Student-t distribution with $n-1$ degrees of freedom, finding a value $t_{n-1,\alpha/2}$ satisfying

$$P(-t_{n-1,\alpha/2} \leq T_{n-1} \leq t_{n-1,\alpha/2}) = 1 - \alpha$$

The mean of a normal population with unknown variance

Procedure

- The following interval corresponds to the required probability

$$-t_{n-1,\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{n-1,\alpha/2}$$

- Replacing sample values and solving for the value of μ in the inequalities we obtain the desired confidence interval

$$\bar{x} - t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \leq \mu < \bar{x} + t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$$

The mean of a normal population with unknown variance

Exercise 1.6

You have measured the working life of a sample of 20 high-efficiency light bulbs. For this sample the measured mean life has been equal to 4520 h., with a sample standard deviation equal to 750 h. If the working life of these bulbs is assumed to follow a normal distribution, compute a confidence interval at the 95 % level for the average life of all the bulbs (population mean)

Results

$$\alpha = 1 - 0{,}95 = 0{,}05, \qquad t_{n-1,\alpha/2} = t_{19,0{,}025} = 2{,}093$$

$$-2{,}093 \leq \frac{4520 - \mu}{750/\sqrt{20}} \leq 2{,}093$$

$$4169{,}0 \leq \mu \leq 4871{,}0$$

The mean of a normal population with unknown variance

Exercise 1.6

You have measured the working life of a sample of 20 high-efficiency light bulbs. For this sample the measured mean life has been equal to 4520 h., with a sample standard deviation equal to 750 h. If the working life of these bulbs is assumed to follow a normal distribution, compute a confidence interval at the 95 % level for the average life of all the bulbs (population mean)

Results

$$\alpha = 1 - 0{,}95 = 0{,}05, \qquad t_{n-1,\alpha/2} = t_{19,0{,}025} = 2{,}093$$

$$-2{,}093 \leq \frac{4520 - \mu}{750/\sqrt{20}} \leq 2{,}093$$

$$4169{,}0 \leq \mu \leq 4871{,}0$$

Variance of a normal population

Motivation

- Up to now we have only considered confidence intervals for the population mean
- In some cases we may also be interested in knowing confidence intervals for the variance
- The relevant distributions are not known in general, except for a few cases
- We will only consider the case of normal data

Hypotheses and goal

- We assume that
  - we have a simple random sample of size n
  - the population follows a normal distribution
- Goal: Goal: construct a confidence interval for the (unknown) population variance σ^2
  - For a confidence level $1-\alpha$, either prespecified or selected by us

Variance of a normal population

Procedure

- In this case the basic result is

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

  where S denotes the sample standard deviation and χ^2_{n-1} denotes the chi-squared distribution with $n-1$ degrees of freedom
  - It is an asymmetric distribution that takes nonnegative values
- As in the preceding cases, the first step is to construct an interval that has the desired probability under the chi-squared distribution
  - As the chi-squared distribution is asymmetric, we need two values to define the interval, $\chi^2_{n-1,1-\alpha/2}$ and $\chi^2_{n-1,\alpha/2}$

Variance of a normal population

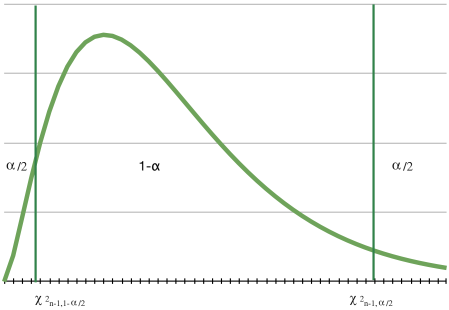
Procedure

- We select these values from the conditions

$$P(\chi^2_{n-1} \geq \chi^2_{n-1,1-\alpha/2}) = 1 - \alpha/2, \qquad P(\chi^2_{n-1} \geq \chi^2_{n-1,\alpha/2}) = \alpha/2$$

- They satisfy

$$P(\chi^2_{n-1,1-\alpha/2} \leq \chi^2_{n-1} \leq \chi^2_{n-1,\alpha/2}) = 1 - \alpha$$



Variance of a normal population

Procedure

- The following interval corresponds to the desired probability

$$\chi^2_{n-1,1-\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1,\alpha/2}$$

- Replacing sample values and solving for σ^2 in the inequalities we obtain the interval

$$\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}$$

- For the standard deviation the corresponding confidence interval will be given by

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}}$$

Variance of a normal population

Exercise 1.7

For the data in exercise 1.6, you are asked to compute a confidence interval at the 95 % level for the (population) standard deviation of the bulb life

Results

$$\alpha = 1 - 0{,}95 = 0{,}05$$
$$\chi^2_{n-1,1-\alpha/2} = \chi^2_{19,0{,}975} = 8{,}907$$
$$\chi^2_{n-1,\alpha/2} = \chi^2_{19,0{,}025} = 32{,}852$$
$$8{,}907 \leq \frac{19 \times 750^2}{\sigma^2} \leq 32{,}852$$
$$325323 \leq \sigma^2 \leq 1199899$$
$$570{,}37 \leq \sigma \leq 1095{,}40$$

Variance of a normal population

Exercise 1.7

For the data in exercise 1.6, you are asked to compute a confidence interval at the 95 % level for the (population) standard deviation of the bulb life

Results

$$
\begin{aligned}
\alpha &= 1 - 0{,}95 = 0{,}05 \\
\chi^2_{n-1,1-\alpha/2} &= \chi^2_{19,0{,}975} = 8{,}907 \\
\chi^2_{n-1,\alpha/2} &= \chi^2_{19,0{,}025} = 32{,}852 \\
& 8{,}907 \leq \frac{19 \times 750^2}{\sigma^2} \leq 32{,}852 \\
& 325323 \leq \sigma^2 \leq 1199899 \\
& 570{,}37 \leq \sigma \leq 1095{,}40
\end{aligned}
$$

Confidence intervals

Summary for one population

- For a simple random sample

Parameter	Hipotheses	Distribution	Interval
Mean	Normal data Known variance	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\mu \in \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
	Nonnormal data Large sample	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$	$\mu \in \left[\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$
	Proportions Large sample	$\frac{\hat{P}-p}{\sqrt{\hat{P}(1-\hat{P})/n}} \sim N(0,1)$	$p \in \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$
	Normal data Unknown var.	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$	$\mu \in \left[\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right]$
Variance	Normal data	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$	$\sigma^2 \in \left[\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right]$